Gender Growth Gaps Across Indian States

Pinelopi Goldberg  
*Yale University*, penny.goldberg@yale.edu

Somik Lall  
*World Bank Group*

Meet Mehta  
*Yale University*, meet.mehta@yale.edu

Michael Peters  
*Yale University*, michael.peters@yale.edu

Aishwarya Ratan  
*Yale University*, aishwarya.ratan@yale.edu

Follow this and additional works at: [https://elischolar.library.yale.edu/egcenter-discussion-paper-series](https://elischolar.library.yale.edu/egcenter-discussion-paper-series)

Part of the [Growth and Development Commons](https://elischolar.library.yale.edu/egcenter-discussion-paper-series)

**Recommended Citation**

Gender Growth Gaps Across Indian States

Penny Goldberg, Somik Lall, Meet Mehta,
Michael Peters, and Aishwarya Ratan

June 2024
Preliminary

Abstract

The extent to which women participate in the labor market and have access to formal employment differs greatly across Indian states. In this paper we build on the methodology developed by Hsieh et al. (2019) to estimate the productivity consequences of such differences. Using rich microdata on occupational sorting and earnings, our theory allows to separately identify labor demand distortions (e.g., discrimination in hiring for formal jobs) from labor supply distortions (e.g., frictions that discourage women’s labor force participation. We find that demand distortions are negatively related to state-level economic development, while supply distortions are unrelated to state-level GDP. Equalizing distortions across Indian states could raise state-level productivity by up to 10%.

1 Theory

We consider a simplified version of Hsieh et al. (2019) where we have two groups - men and women, and three occupational choices - employment in formal work ($f$), self employment (informal) ($i$) and home production ($h$). Men and women differ in three dimensions. First, women might face a demand distortion, which we model as an exogenous tax on their labor earnings. Second women might face a supply distortion, whereby choosing a particular occupation reduces their utility. Third, men and women can differ in their occupation-specific
human capital. Crucially, both demand and supply distortions vary across Indian states.

Formally, we model the utility of an individual of group \( g = m, f \), i.e. male or female, in state \( s \) who chooses occupation \( o \) as

\[
\log U_{og}(s) = \log C_{og}(s) + \log z_{og}(s)
\]

where \( C_{og}(s) \) denotes consumption and \( z_{og}(s) \) denotes the utility of working in occupation \( o \). Consumption is linked to individual’s human capital, the prevailing wage rate (per efficiency unit), and the prevailing demand distortion by the budget constraint

\[
C_{og}(s) = (1 - \tau_{og}(s)) w_o(s) \bar{h}_{og} \epsilon.
\]

Here \( w_o(s) \) is the prevailing wage rate in occupation \( o \) in state \( s \), \( \bar{h}_{og} \) denotes the occupation specific human capital of an individual of group \( g \), \( \tau_{og}(s) \) parametrizes the demand distortion and \( \epsilon \) is an idiosyncratic productivity draw that allows individuals to differ in their comparative and absolute advantage in different occupations. Note that the human capital term \( \bar{h}_{og} \) differs across occupations and groups (i.e., we allow for male and female skills to differ across occupations in an unrestricted way) but that they are not allowed to differ across states \( s \). As discussed later, this will form a key identification assumption for us going forward.

Substituting (2) into (1), indirect utility is given by

\[
U^*_og(s) = \tilde{w}_{og}(s) \epsilon_o
\]

where

\[
\tilde{w}_{og}(s) \equiv w_o(s)(1 - \tau_{og}(s)) \bar{h}_{og} z_{og}(s)
\]

Hence, \( \tilde{w}_{og}(s) \) summarizes the systematic attractiveness of an occupation \( o \) in state \( s \) for group \( g \) and it depends on occupational skill prices \( w_o(s) \), labor demand distortions \( \tau_{og}(s) \), the average human capital endowment \( \bar{h}_{og} \), and labor supply distortions \( z_{og}(s) \).

Individuals choose occupation \( o \) to maximize \( U^*_og(s) \). For tractability we assume that \( \epsilon_o \) is drawn from a multivariate Frechet distribution:

\[
F_g(\epsilon_f, \epsilon_i, \epsilon_h) = \exp(-\epsilon_f^{-\theta} - \epsilon_i^{-\theta} - \epsilon_h^{-\theta})
\]

This leads to our main result:

**Proposition 1** (Occupational Sorting and Average Wages): Let \( p_{og}(s) \) denote the fraction of people from state \( s \) and group \( g \) who choose occupation
o and $\bar{w}_{og}(s)$ denote the geometric average of earnings in occupation $o$ by state $s$ of group $g$. Then

$$p_{og}(s) = \frac{\tilde{w}^\theta_{og}}{\sum_j \tilde{w}^\theta_{jg}} \tag{3}$$

$$\bar{w}_{og}(s) = \tilde{\Gamma} \left( \sum_j \tilde{w}^\theta_{jg} \right)^{\frac{1}{\theta}} z_{og}(s)^{-\frac{1}{\theta}} \tag{4}$$

Equations (3) and (4) are at the heart of our identification strategy. In particular, using (4) to substitute $\bar{w}_{og}(s)$ for the labor supply distortions $z_{og}(s)$ yields that the fraction of women working in an occupation $o$ (relative to men) is given by

$$p_{ow}(s) = \frac{p_{om}(s)}{\bar{w}_{ow}(s)} \left( 1 - \tau_{ow}(s) \right)^\theta \times \left[ \frac{\bar{h}_{ow}}{\bar{h}_{om}} \right]^\theta \times \left[ \frac{\bar{w}_{ow}(s)}{\bar{w}_{om}(s)} \right]^{-\theta} \tag{5}$$

Equation (5) highlights that relative occupational shares are driven by three considerations: women can be underrepresented in a particular occupation if (i) they face a labor demand distortion, (ii) they have a lower human capital endowment in this occupation and (iii) their average wage is relatively high. This effect of the wage works through the labor supply distortion. Recall that equation (4) highlights that average occupation wages only reflect the labor supply distortion $z_{og}(s)$ (note that all other terms are constant across occupations for a given group). A high wages in a particular occupation thus reflects a compensating differential for an existing labor supply distortion. Hence, this reduces the average wage.

**Firms:** A representative firm in state $s$ produces final output as follows:

$$Y(s) = \left[ \sum_o (A_o(s)H_o(s))^{\frac{\sigma - 1}{\sigma}} \right]^{\frac{\sigma}{\sigma - 1}} \tag{6}$$

where $H_o(s) \equiv \sum_g q_g(s)p_{og}(s)E[h_{og}(s)e_{og}]$

Total output of the country is given by

$$Y = \sum_s Y(s)$$
Determinants of Labor Market and Human Capital Frictions:
Firms have “taste” for discrimination of women workers. The utility of firm owner in state $s$ is given by:

$$U_{owner}(s) = Y(s) - \sum_{o} \sum_{g} (1 - \tau_{og}^w(s)) w_o(s) H_{og}(s) - \sum_{o} \sum_{g} d_{og}(s) H_{og}$$ (B1)

Under perfect competition, we have $\tau_{og}^w(s) = d_{og}(s)$. Therefore, frictions are ultimately pinned down by discriminatory tastes of the owners.

Equilibrium
Equilibrium in this economy is solved for each state $s$ separately. A competitive equilibrium consists of sequence of occupational choices, total efficiency units of labor in each group $H_{og}$, final output $Y$, and an efficiency wage $w_o$ in each occupation such that

1. Each individual chooses occupation that maximises lifetime utility: $o^* = \arg \max_o U(\tau_{og}^w, z_{og}, w_o, \epsilon_o, \mu_o)$ taking as given $\{\tau_{og}^w, z_{og}, w_o, \epsilon_o, \mu_o\}$.
2. Perfect competition in the final goods sector generates $\tau_{og}^w(s) = \frac{d_{og}(s)}{w_o(s)}$. Therefore, frictions are ultimately pinned down by discriminatory tastes of the owners.

Proposition 6 (Solving for the general equilibrium): The general equilibrium of the model is $H_{og}^{supply}, H_{o}^{demand}, w_o$ and market output $Y$ for each state $s$ such that

1. $H_{og}^{supply}(s)$ aggregated the individual choices:
   $$H_{og}^{supply}(s) = q_g(s)p_{og}(s)E[h_{og}(s)\epsilon_{og}(s)|choose o]$$
   where $q_g(s)$ denotes number of workers of group $g$ and state $s$.
2. $H_{o}^{demand}(s)$ satisfies firm profit maximisation:
   $$H_{og}^{demand}(s) = \left(\frac{A_o(s)\frac{\sigma}{\sigma-1}}{w_o(s)}\right)^\sigma Y(s)$$
3. $w_o(s)$ clears each occupational labor market:
   $$H_{om}^{supply}(s) + H_{ow}^{supply}(s) = H_{o}^{demand}(s)$$
4. Total output is given by production function in equation 7 and equals aggregate wages + total revenues from $\tau_{og}$. 

4
2 Identification

From data, we will have the estimates of $p_{ig}(s)$ and $\bar{\text{wage}}_{og}(s)$. We use the assumption on $\theta$ and $\bar{h}_{og}$ to calibrate the parameters $\tau_{ow}(s)$ and $z_{og}(s)$.

2.1 Labor Market Distortions: $\tau_{og}(s)$

We estimate $\tau_{ow}(s)$ in two steps. First, we estimate the value of net taxes $(1 - \tau_{ow}(s))$ for state $s$, relative to a reference state $k$. From equation 5, using the identifying assumption that $\bar{h}_{ow}/\bar{h}_{om}$ is same across states, we get

$$\frac{1 - \tau_{ow}(s)}{1 - \tau_{ow}(k)} = \left[\frac{p_{ow}(s)/p_{om}(s)}{p_{ow}(k)/p_{om}(k)}\right]^{\frac{1}{\theta}} \times \left[\frac{\bar{\text{wage}}_{ow}(s)/\bar{\text{wage}}_{om}(s)}{\bar{\text{wage}}_{ow}(k)/\bar{\text{wage}}_{om}(k)}\right]$$

(7)

Next, we assume that the average tax on women for occupation $o$ equals to the number $\tau_{ow}$ which equals:

$$\tau_{ow} = \sum_s \frac{L(s)}{L} \tau_{ow}(s)$$

where $L(s)$ is population in state $s$ and $L$ is total population in the country. Given the equation above, since $\sum_s \frac{L(s)}{L} = 1$, we have

$$1 - \tau_{ow} = \sum_s \frac{L(s)}{L} (1 - \tau_{ow}(s))$$

Dividing the above equation by $(1 - \tau_{ow}(k))$, we get:

$$1 - \tau_{ow}(k) = \frac{1 - \tau_{ow}}{\sum_s \frac{L(s)}{L} \left(\frac{1 - \tau_{ow}(s)}{1 - \tau_{ow}(k)}\right)}$$

where in the R.H.S., we use the relative net taxes we obtained from equation 7.

Once we get the value of $(1 - \tau_{ow}(k))$, it is straightforward to get the value of $\tau_{ow}(s)$ using the relative net taxes.

2.2 Norms: $z_{og}(s)$

For $z_{og}(s)$, we use the relative wages from equation ???. Notice that we have the normalisation $\bar{z}_{hg} = 1$ for both the groups. This gives us the following equation:
\[
\frac{\text{wage}_{og}}{\text{wage}_{hg}} = z_{og}^{-1}
\] (8)

This normalisation can also be used to estimate \( m_g(c) \). Rearranging equation (8) for the home sector, we get:

\[
\hat{m}_g(c) = \bar{\text{wage}}_{hg}(c, c)^\theta \tilde{\Gamma}^{-\theta}
\]

**2.3 \( \theta \)**

We will take the value of \( \theta = 1.36 \), which is the value of \( \theta(1 - \eta) \) in the paper. This is chosen since we abstract away from human capital choice and hence, \( \eta = 0 \).

**2.4 Imputing home sector wages**

Note that equation 5 requires us to know the wages for the home sector for both men and women. For men, we assume that whenever they decide not to work for formal or informal sector, they can always go back and work as a casual labor in agriculture sector. Therefore, we set the wage (earnings) in the home sector for men to be equal to wage (earnings) as a casual labor in the agriculture sector.

For the wages of women in the home sector, we can proceed in two ways. One way is to set it equal to the wage of casual female labour in the agriculture sector. The other way is the exploit the assumption that \( \tau_{hw} = 0 \). Then use equation 5 to estimate wages relative to reference state \( k \). This is can only help us in getting relative preferences and not the levels. For the time being, we proceed with the first approach.

**2.5 \( A_0(s), w_0(s), Y(s) \)**

To estimate \( A_0(s) \), we use the labor demand equation in point 2 of proposition 6 above. To use that equation, we additionally require the estimates of \( H_{og}^{\text{demand}}(s), Y(s) \) and \( w_0(s) \). For \( w_0(s) \), we use equation (??), which can be re-written as follows for the male group (given \( h_m = 1 \) and \( \tau_{om}^w = 0 \)):

\[
w_0(s) = \left( p_{om}(s) \hat{m}_m(s) \right)^{1/\sigma} \left( \tilde{z}_{om}(s) \right)
\]

where has used the estimate of \( \hat{m}_m(s) \) derived while estimating \( z_{og}(s) \).
Given the estimate of $w_o(s)$, we can find labor supply:

$$H^\text{supply}_o(s) = \sum_g q_g(s)p_{og}(s)E[h_{og}(s)e_{og}(s)|\text{choose } o]$$

$$H^\text{supply}_o(s) = \sum_g q_g(s)p_{og}(s)\frac{\theta_0}{\theta_0 - 1} h_{og}(s) \theta(1 - \frac{1}{\theta_0})$$

We use the equilibrium relationship $H^\text{supply}_o(s) = H^\text{demand}_o(s) = H^o(s)$ to plug back into the labor demand equation and finally, we can estimate $Y$ as a sum of total wage payments and taxes:

$$Y(s) = \sum_o w_o(s)H^o(s)$$

Lastly, following the paper, we pick $\sigma = 3$ arbitrarily. Detailed derivation of above is provided in handwritten appendix.

### 2.6 Implied $h_{ow}$ and $h_{om}$

Note that equation 7 provides us with relative tax rates for different states. It is only the level of these tax rates which is pinned down by the choice of average tax rate $\tau_{ow}$. Note that the choice of average tax rate is equivalent to making a choice about $h_{ow} \over h_{om}$ since it is invariant across states. Therefore, once we have recovered the level of tax rates across states, we can use equation 5 for any of the states to get the values of implied $h_{ow} \over h_{om}$. After normalising $h_{om} = 1$ in all the occupations, we can use the values of $h_{ow}$ for computing the sectoral level human capital supply while solving for the equilibrium. It is important to note that the implied levels of human capital is only used for computing the counterfactual equilibrium, and not directly used for computing any of the parameters.

### 2.7 Arithmetic and Geometric Mean

Lastly, we observe arithmetic average of wages in the data where equation 5 and ?? requires geometric average of wages. We convert arithmetic average wage to geometric average wages using the following formula, which is derived in the handwritten appendix.

$$\text{wage}_{og} = \text{Aritmetic Average Wage}_{og} \frac{\tilde{\Gamma}}{\Gamma(1 - \frac{1}{\theta_0})}$$

where $\tilde{\Gamma} = e^{\gamma_{em} \theta_0}$. 

7
3 Solving for Equilibrium

Given the parameters, we can now move to solve the model. The way to solve for equilibrium here is to guess the values of $Y(s)$, $m_m(s)$ and $m_w(s)$. These 3 guess, along with the model parameters help us in calculating $H_o^{\text{demand}}(s)$ and $H_o^{\text{supply}}(s)$. We find $w_o(s)$ such that the two are equal and re-estimate the values of $Y(s)$, $m_m(s)$ and $m_w(s)$. We iterate this process until the guess and the estimated values of $Y(s)$, $m_m(s)$ and $m_w(s)$ are equal.

4 Misallocation Across Indian States

4.1 Data

We apply this model to India using the Periodic Labour Force Survey (PLFS) of 2018-19. It contains detailed data on labor force participation of a representative sample of (approx.) 100,000 households. PLFS classifies the employment status of an individual into following categories: Regular salaried employee (code 31, 71, 72), Self employed (code 11, 12, 21, 61, 62), Casual wage labour (code 41, 51), Unemployed (code 81) and Out of labour force (code 91-97). We classify regular salaried employee into the formal sector ($f$), self employed and casual wage labour into informal sector ($i$), and unemployed and out of labour force in the home sector ($h$). We apply standard data cleaning procedures.\(^1\)

4.2 Estimated Distortions

From PLFS, we get the data on $p_{og}(s)$ and $\bar{\omega}_{og}(s)$. We plug these into equations 7 and 8 to estimate the distribution of labor taxes $\tau_{ow}(s)$ and supply distortions $z_{ow}$. We assume $\theta = 1.5^2$ and $\tau_{iw} = \tau_{fw} = 0.7$. These assumptions imply that the human capital ratios for our model are $\frac{h_{iw}}{h_{im}} = 0.69$ and $\frac{h_{fw}}{h_{fm}} = 0.98$.

---

\(^1\) We restrict the sample to working age population: age 25-60 years. Further, we drop states where any of the state-sector-group cells have less than 50 observations. We winsorize the wage data in every state-sector cell at the 5\(^{th}\) and 95\(^{th}\) percentile. Lastly, for individuals who work as unpaid labour in the informal sector, we assign their wage equal to the average wage of other informal sector employee in their relevant state-group cell.

\(^2\) Hsieh et al. (2019) estimates $\theta = 1.52$ after adjusting for the elasticity of human capital w.r.t. human capital expenditure. We assume $\theta = 1.5$ as of now, which can be further updated by fitting the Frechet distribution on the wage data.
Figure 1: Labor taxes and supply distortions against state GDP per capita

Figure 1 plots the estimated labor taxes and supply distortions as a function of state level GDP (in logs). The first row plots the labor taxes for the informal \((i)\) and formal \((f)\) sector respectively. We see wide variations in level of labor taxes across states, and more importantly, we note that labor taxes decline on average with an increase in state GDP per capita.

The second row in figure 1 plots the ratio \(\frac{z_{iW}(s)}{z_{fW}(s)}\) for the informal \((i)\) and formal \((f)\) sector respectively. For the formal sector, we see that this ratio is below 1 for most of the states. This implies that the estimated supply distortions for the formal sector is relatively more for women compared to men in most of the states. For the informal sector, the variation in relative distortions across states is relatively lower compared to the formal sector and we see roughly half of the states with ratio above and below 1. We interpret this as labor supply distortions being more severe for women in the formal sector compared to the informal sector.

### 4.3 Implications for Productivity

Distortions on either supply or demand reduce productivity through a misallocation of talent - more productive women end up working in less productive sector. We now quantify the economic costs of our estimates from Section 4.2 for productivity both at the aggregate (i.e. India-wide) level and across states.

We study two counterfactual scenario: 1. Women face no labor demand
Figure 2: Counterfactual GDP and Employment Choices of Women

distortions relative to men (τ_{iw} = τ_{fw} = 0). Women face no labor supply
distortion relative to men (z_{ow} = z_{om} for o ∈ \{h, i, f\}).

In Figure 2 we focus on the aggregate level. Specifically, we report the changes in aggregate GDP and women’s employment choices in the two counterfactual cases. Removing labor demand distortions relative to men leads to higher increases in GDP (6.5 p.p.) compared to removing labor supply distortions relative to men (1.9 p.p.). These positive effects on aggregate productivity are due to changes in occupational sorting. In particular, removal of labor demand distortions lead to significant reallocation of women from the home sector to the informal and formal sector. Female labor force participation rate increases more than 37 p.p. on aggregate, with more women entering the informal sector compared to formal sector. When labor supply distortions are removed, there is a small impact on the labor force participation of women (3.2 p.p. increase). However the allocation towards formal sector improves compared to the informal sector.

The analysis in Section 4.2 suggests that distortions, especially on the labor demand side, are much more prevalent in some states relative to others. In Figure 3 we therefore estimate such counterfactual gains for each Indian state. In the first plot, we focus on demand distortions \(\tau_{og}\). We find that there is substantial heterogeneity across states. Relatively poor states such as Bihar and Uttar Pradesh could increase their GDP by up to 11.3 % and 8.5 % respectively if women would not face a distortions on the labor market which
increases their marginal product above the ones of men. By contrast, such growth potential is much lower in rich states such as Kerala and Tamil Nadu, where differences in implicit “labor taxes” are less pronounced. In the second plot of figure 3, we focus on supply distortions $z_{og}$. We find that the productivity losses due to supply distortions are substantially smaller and their relation with GDP per capita is weaker compared to the first plot. If anything, the gains from removal of labor supply distortions appear to be weakly increasing across the states.

5 Conclusion

References

6 Appendix